INSTITUTO FIGUEROLA
DE HISTORIA Y CIENCIAS SOCIALES
Universidad Carlos III de Madrid

Universidad Carlos III de Madrid

Working Papers in Economic History

April 2016 WP 16-03

Knowledge Shocks Diffusion and the Resilience of Regional Inequality

Alexandra López-Cermeño

Abstract
This paper provides a simplified method of exploring the geographical limits of a knowledge shock over the long run. Using a geographically decomposable distance weighed sum of world GDPs by county, differences in differences regression analysis shows that a new university will not only have a positive impact on the local economy, but also on the GDP of nearby counties. Furthermore, challenging the conventional wisdom that knowledge spillovers affect the local economy, this study provides evidence that the effect expands to the whole nation although its strength dilutes with distance. Consistent with the education literature, this investigation provides evidence that the shock will make the relative GDP of foreign competitors worse-off. Results are persistent in the long run, although the effect of time is also decreasing. Results are robust to potential endogeneity related to the self-selection of prosperous allocations for new academic institutions.

Keywords: New Economic Geography, Spillovers, U.S. Counties.

JEL Classification: L8, N72, R11, O18

Alexandra López-Cermeño: Departamento de Ciencias Sociales, Universidad Carlos III, Calle Madrid, 126, 28903 Getafe, Spain.
E-mail: alexandra.lopez.cermeno@uc3m.es

UNIVERSIDAD CARLOS III DE MADRID • c/ Madrid 126 • 28903 Getafe (Spain)• Tel: (34) 91 624 97 97
Site: http://portal.uc3m.es/portal/page/portal/instituto_figuerola/home/publications/working_papers

Publisher:
Carlos III University of Madrid. Figuerola Institute of Social Sciences History
www.uc3m.es/if

Series:
Working Papers in Economic History
ISSN: 2341-2542

Electronic version of these working paper series available on:
http://hdl.handle.net/10016/16

Knowledge Shocks Diffusion and the Resilience of Regional Inequality

Alexandra López-Cermeño*

Universidad Carlos III de Madrid

e-mail: allopezc@clio.uc3m.es

2016

An investment in knowledge always pays the best interest.

Benjamin Franklin [1758]

Abstract

This paper provides a simplified method of exploring the geographical limits of a knowledge shock over the long run. Using a geographically decomposable distance-weighed sum of world GDPs by county, differences in differences regression analysis shows that a new university will not only have a positive impact on the local economy, but also on the GDP of nearby counties. Furthermore, challenging the conventional wisdom that knowledge spillovers affect the local economy, this study provides evidence that the effect expands to the whole nation although its strength dilutes with distance. Consistent with the education literature, this investigation provides evidence that the shock will make the relative GDP of foreign competitors worse-off. Results are persistent in the long run, although the effect of time is also decreasing. Results are robust to potential endogeneity related to the self-selection of prosperous allocations for new academic institutions.

JEL classification: L8, N72, R11, O18.

Keywords: New Economic Geography, Spillovers, U.S. Counties.

*I am grateful to the participants of the 10th SOUND Economic History Workshop for all their useful comments and questions and to Universidad Carlos III de Madrid for providing funds to perform this research project.

1 Introduction

The search for policies to fight the regional persistence of inequality is crucial in industrialized economies. A number of contributions to the economic history literature have shown that high-value added sectors tend to cluster in particular regions and promote a process of de-industrialization in the rest of the economy, creating long-term divergences that lead the population to move toward these clusters in search of higher income.[1] These policies might include national subsidies, tax reductions, federal minimum wage increases or locally planned projects that promote business creation or service provision like new airports, freeways or the improvement of local administration to hopefully increase or maintain the population (Moretti [2012]).

In 1949, the Federal Government of the United States decided to investigate the power of nuclear energy. The fear that nuclear power could harm the health of the population led to the search for an isolated desert to locate the nuclear energy research facility.[2] In less than two decades, Idaho, formerly known as the *Potato State*, was among the top-100 biggest metropolitan areas in the country. The population of Idaho Falls and its surrounding counties had increased, had become much more skilled, and enjoyed higher living standards. The creation of this national research facility led to a presumably unexpected upswing in terms of population and income that accumulated further growth in nearby counties. This study explores the geographical impact of a knowledge shock as urged by two fathers of the New Economic Geography (Fujita and Krugman [2003]) and challenges the conventional wisdom that knowledge spillovers act locally. Evidence shows that the effect of local investments tend to spread to the whole nation and make foreign competitors comparatively worse-off, *ceteris paribus*. These conclusions provide optimistic prospects for local policy-makers that can contest regional inequality, but also promote national competitiveness through micro-investments.

I propose an experiment that considers the appearance of a new academic institution as a knowledge shock and performs a common differences in differences methodology to observe the significance of the knowledge shock in comparison with an untreated control group. The main contribution of this paper is the simplification of the methodology to obtain results that would usually require complicated spatial econometrics: by disentangling the county geographical impact of GDP into layers, I can test the significance of the shock at different levels, avoiding the nuisance of complicated county neighboring matrices. An additional

[1] Enflo and Rosés [2015] explore the Swedish late industrialization period and the policy efforts to decrease regional inequality, Autor et al. [2008] do the same for the US in the last few decades.

[2] National Reactor Testing Station.

contribution is the consideration of the whole range of counties in the USA rather than only cities or metropolitan areas, which usually lead to biased conclusions.

The basic empirical results show that the establishment of new universities during the 20^{th} century had a positive impact not only on the local economy and its nearby counties, but also on distant locations within the nation. Moreover, *ceteris paribus*, a new academic institution in any county of the USA made foreign competitors relatively worse-off in terms of GDP. The effects of knowledge shocks seem stronger in closer locations and milder, but significant, in more distant areas. Similarly, the effect of the shock seems to slowly wear-out over time. However, these effects seem to persist over the whole century. Testing the significance of the shock in per capita terms shows that the effect in productivity is only local and the shock affects nearby regions through a multiplier effect as claimed by Moretti [2010].

The remainder of the paper is organized as follows. Section 2 discusses the historical background on the higher education system of the United States in context with urban growth during the last century. Section 3 develops the theoretical framework and main hypothesis, followed by a description of the variables and the empirical strategy. Section 5 presents results and robustness checks and Section 6 concludes.

2 Historical background

The economic history of the United States is a story of skills and human capital. Its academic institutions have not only turned the Human Capital Century into the American Century (Goldin and Katz [2009]), but have also driven the divergence of regional economic performance. While the relevance of academic and research institutions has only recently become evident, it has been an important driver of economic growth for a long period of time. The following paragraphs summarize the origins of the American higher education system.

As a consequence of the fear that the imprudent European tendencies would corrupt their souls, the Puritans who travelled to the New World launched the precursors of college institutions in the first settlement allocations at the end of the 17^{th} century. These would become the well-known institutions of Harvard, Yale, and William & Mary Universities. Originally these were meant to produce educated gentlemen whose "business (was) to spread religion and learning among mankind" [Geiger, 2014, pp. 11]. The evolution and emergence of these institutions was slow and always related to local religious elites.

The initial courses included theological and literary education and grammar subjects; the introduction of ancient languages like Greek and Hebrew to study sacred original texts motivated the introduction of logic and other mathematical areas, but the low depth of scientific knowledge did not yet reflect intellectual advances in Europe until well into the 18th century. Eventually, the American elite accepted that scientific knowledge could make laborers more productive, and thus Newtonian scientific doctrine started to be taught in US colleges. By 1836 the academic system even allowed higher education for women. However, the elitist character of these institutions forced the imbalance between theoretical science and practical applications for gentlemen. Under the rising scarcity of mechanical and agricultural engineers, fostered by the railroad boom of the 1840s, some non-college alternatives like mechanics and polytechnic institutes started to arise in the cities.

The role of private sector investment went hand in hand with the growth of "useful knowledge". By the second half of the 19th century, America had turned into a world reference of technological advance. The land grant promoted by President Lincoln (Morrill Act, 1862) helped revolutionize higher education by providing states with public lands to create universities specializing in agriculture, mechanics, and military tactics. This was the beginning of mass higher education, as congressman Morrill envisioned the existence of a college in each state as an opportunity "*accessible to all, but especially to the sons of the toil . . . thousand willing and expecting to work their way through the world by the sweat of their brow*", [Geiger, 2014, pp. 281].

In the 20th century, the higher education system offered students the widest range of opportunities in the world (around 1,400 institutions offering bachelor's degrees in several areas). As Goldin and Katz [2009] put it, the system was geographically spread and accessible to all kinds of economic and intellectual backgrounds. While the role of mass education has been crucial for the productive structure of the country, the location of Universities seems to create regional divergences in the US territory. First, colleges were established even when secondary education was not yet standardized, so that the mass movement toward college had to be led by the diffusion of secondary schooling. Although both kinds of institutions were originally decentralized, public and open to all genders and races, a minimum scale was needed to create such institutions. Just as the first colleges appeared in the first populated settlements, secondary schools were allocated in towns with at least 3,000 people in 1903 (Goldin and Katz [2009]). This threshold set a precedent for the divergence between rural and urban growth.

The effects of academic education on growth are directly visible on labor productivity through an increase in the quality of the workforce (see Caselli and Coleman [2002]), but there are also indirect effects: higher income generated by labor productivity raises physical capital investment and the capital to labor ratio; also, the quality of the workforce facilitates the diffusion of innovations and ideas. Thus, regions with higher levels of education are expected to grow faster. At the beginning of the century, all the regions below the population boundary would be at a practical disadvantage of growth opportunities, although evidence suggests that town size is negatively correlated with school attendance rates [Goldin and Katz, 2009, pp. 224]. This negative correlation is explained by the relative opportunity costs of schooling as well as the quality derived by a higher share of students per school in biggest cities. Nevertheless, the demand for skills gradually changed in the cities. US occupational data from Edwards [1943] show that in the late 19th century, only 10 per cent of the workforce was engaged in jobs requiring education beyond secondary school, by 1920 more than a quarter of the jobs required high school or college education. The proliferation of white-collar occupations was accompanied by the structural change of the economy.

Regional specialization determines the average level of human capital: while mining regions are associated with relatively lower effects of knowledge spillovers and have remained small, cities that grew around the textile industry were crowded with unskilled labor and only grew at the beginning of the century. In contrast, commercial towns that specialized in skill-intensive activities like accounting, advertising and law tended to become large cities over the same period (Chicago, Boston, New York). This way, in the 1930s the population in Idaho Falls was specialized in the production of agricultural products and their low wages responded to their skills level. After the knowledge shock, represented by the establishment of the Nuclear Research Center, their production bundle diversified by including valuable knowledge intensive services (nuclear energy research), and the population grew by attracting scientific employees that earned much higher salaries, fostering the creation of new businesses and, eventually, raising living standards.

The analogy between the Human Capital Century and the American Century is not only motivated by the higher human capital increase of the American labor force over the century, but also because the change was not comparable to the standards of any other nation, which led to the great divergence between the US and the rest of the world and also the increasing domestic inequality within the country. The next section explores the different views proposed in the literature for exploring the role of educational differentials on regional disparities, and explains how this investigation differs from spatial models to account for knowledge shocks across space and time.

3 Theoretical framework

The link between urban growth and human capital has been widely studied. Glaeser et al. [2014, 1995] and Simon and Nardinelli [2002], among others, identified human capital and skills as an important factor behind the growth of cities after WW2. The theoretical base comes from the evidence that the existence of urban clusters is derived from the positive external effects of human capital (increasing returns to scale); without these, rational citizens would not bear the costs of moving to crowded clusters just as observed by Lucas [1988]. Endogenous growth models argue that innovation comes from the mix of labor, human capital and knowledge. In this sense, the proliferation of institutions providing human capital is expected to promote higher growth and urbanization. A crucial question is whether urbanization is exogenous to the localization of new academic institutions or whether there is a self-selection process instead.

This paper offers a view in which the localization of a new academic institution acts as a positive knowledge shock to the county. The objective is not to show that a shock improves the local economy, as that has already been repeatedly proven on several occasions (Anselin et al. [1997],Goldin and Katz [2009], and Krueger and Lindahl [2001]). Instead, this research addresses the extent to which local spillovers spread geographically and whether these effects persist over time. The analysis of such shocks on economic activity is not new in economic history, however the spatial diffusion of local shocks over the long run has never been investigated with this much detail. In fact, two pioneers of the New Economic Geography literature recently advocated for the need for this kind of analysis (Fujita and Krugman [2003]).

The literature on urban economics has explored the effects of external shocks on the spatial distribution of economic activity in several areas, finding mixed results. The pessimistic view can be well exemplified by Redding et al. [2011], who use the German division after WW2 and its later reunification as a natural experiment of an external shock to the location of the air transport industry. They show that neither endowments nor market access differentials are big enough to explain the reallocation of the air hub from Berlin to Frankfurt. They suggest that the differential between local economic activity is not a good predictor either, instead, the selection of Frankfurt as the localization of the main air hub in Germany responds to a relatively small external intervention (US setting Frankfurt as the main air transport base) that influenced the location of the new hub given the large investments required for its functioning. Their conclusion that German reunification was not a sufficiently big shock to return the hub to its pre-war location casts evidence against the ability of policy to shift economic activity from an existing steady state. On the same track, Davis and Weinstein [2002] showed that the allocation of cities in Japan was persistent over 8 thousand years and the massive destruction of the atomic bombs did not alter the original allocation of the main cities. This localization persistence is explained by fundamentals and the degree of inequality is accounted for by increasing returns, but again, the effect of shocks is only temporary and does not change the steady state. WW2 bombings in Germany also offer an opportunity to analyze the reconstruction of markets in their original cities. According to Brakman et al. [2004], the effects are significant for both areas; however the bombings affected the reallocation and growth of cities permanently in East Germany and temporarily in West Germany, where increasing returns slowly took over city growth path in the long run.

On the other hand, a number of studies reveal opposing conclusions. Local shocks generated by the expansion of transport facilities like the railroad (Atack et al. [2010]) or highways (Baum-Snow [2007]) have provided evidence on changes in the urbanization rates of counties and cities in the mid-19$^{\text{th}}$ and mid-20$^{\text{th}}$ century. Other authors have proved the significance of negative shocks like wars, or factor input shortages. For example, Hanlon [2014] studies the American cotton supply drop during the Civil War, finding temporary growth effects and permanent level effects on population in British cotton towns.

These contradicting views on the effects of external shocks are generally looking for changes in steady state aggregates or local changes in levels or growth magnitudes in comparison to a prior situation; however, the effects of local shocks on neighboring areas remains largely unexplored. Hornbeck and Keniston [2014] have surveyed this area by performing a very local study on the impact of the Great Fire of Boston in 1872 on land prices of unburned nearby areas. They argue that in a period of intense growth, the fire motivated a reconstruction that increased the property values of the burned areas more than proportionally. This was a consequence of the parallel reformation of nearby buildings, showing that the reconstruction led to the spread of local spillovers. In the same line of work, Simon and Nardinelli [2002] find that the extent of spillovers works at the city level.

Scholars have tried to study the economic impact of research institutions following two paths. Taking a microeconomic perspective, some scholars evaluate single-case institutions through the analysis of economic fundamentals like spending, investment and employment rates or more sophisticated variables like the creation of spin-off firms, or the assessment of university-linkaged firms; others use surveys on firms evaluating the local effects of an institution on their decisions. Using a macroeconomic approach, economists like Grilliches [1979] and Jaffe [1989] have generated models based on knowledge production functions derived from the location of institutions, while others have designed cross-sectional econometric experiments that evaluate the economic impact of these institutions. These four approaches have different benefits and drawbacks and are efficiently summarized in Drucker and Goldstein [2007]. In this perspective, the framework I use follows an econometric cross-section experiment on different years (because the data set is actually a panel).

In this area of work, Anselin [2000], using a knowledge production function *à la* Grilliches-Jaffe, showed that universities generate local spillovers across particularly high technology sectors like Electronics and Instruments, extending up to a 75-mile range to the boundary of Metropolitan Areas, while Drugs and Chemicals and Machinery showed no significant spillover effects. Positive evidence on university R&D was previously suggested at the state level by Anselin et al. [1997] using the same approach. Additional studies have proven that industries where new economic knowledge plays an important role have a higher propensity to cluster together (Audretsch and Feldman [1996]) in regions and also across nations (Ciarli et al. [2012]). These positive results offer an opportunity to policy-makers who wish to design economic policies to promote regional growth under the influence of increasing returns to scale at various geographic levels. However, Audretsch et al. [2012] underline the need to be cautious about the potential crowding out effect of private research activity under the provision of publicly funded research.

This investigation extends empirical evidence on three different aspects: first, the geographical framework covers the whole territory of the USA using smaller geographical areas than States or Metropolitan Areas. More specifically, the use of counties increases the number of observations from 50 or 125 to more than 3,000; in addition, it provides a much better view of the diffusion of local spillovers by including not only cities or metropolitan areas, but also rural counties, whereas the traditional literature is generally biased towards the isolated analysis of urban areas. Secondly, the sample extends evidence to five benchmark years, presenting a longer-term view of results. Lastly, instead of considering the use of a spatial lags model (Anselin [2000]; Anselin et al. [1997]), the methodology explores the spatial extent of spillovers using an alternative measure: a distance-weighed sum of GDP that can be decomposed at different geographical scales to identify the extent of the shock.

The use of this framework allows one to ascertain the impact of a new university over distance and time. Say, a new university is established in Fresno in 1911; by 1915 it will have presumably attracted some students that will eventually become part of a pool of skilled workers. With some luck, the pool of skilled workers can foster the creation of new firms that may capture the knowledge spillovers from the university. The increase in local demand will attract other kinds of workers and services with the consequent rise of local income and wages as explained by Moretti [2004].

The central question is whether this shock will affect the well-being in adjacent counties such as San Benito or even spread to more distant counties like Inyo. Further, will this shock expand to Nevada? Will it maybe affect Kansas? How long will the impact last for any of these layers? The following section outlines the way a distance-weighed GPD index can provide a measure of the effects of a knowledge shock and describes how it has been obtained and provides a description of the variables and their sources. The current data set is a panel on distance-weighed Total GDP observations by county over five benchmark years during the 20$^{\text{th}}$ century (1930, 1950, 1980 and 2010). Thus, the number of observations amounts to a total of 12,512: five time-series observations for 3,130 counties. This measure provides the means to investigate the geographical expansion of the shock econometrically using a simple differences in differences model. The distance-weighed sum of GDPs is similar to the concept of Market Potential, originally defined by Harris [1954] as the sum of the GDPs of potential commercial partners weighed by bilateral transport costs or distances. This specification takes into account that two adjacent counties will have a greater chance of trade than two distant counties, with no need to use a neighboring matrix. In this sense, it can be defined as a location's accessibility to the other markets, and can be formulated as:

$$MP_i = \sum_{(j-1)}^{(j-n)} \frac{M_j}{d_i j} \tag{1}$$

Market Potential accounts not only for local GDP, but also for all potential trade with neighboring counties within a state, with the nation as a whole, and even with other nations; this is useful because it provides a detailed view of the extension of the market based on bilateral transport costs and the size of other markets. More background information on Market Potential can be found in the appendix. In a similar way, the distance-weighted total GDP can be split into different components to address the extension of the impact of a new university.

Distance-weighed Total GDP observations for each county have been constructed following the methodologies used by Crafts and Mulatu [2005] and Martínez-Galarraga [2012], who compute Market Potential using its different geographic components. This data set provides visibility on Foreign, Domestic, State and county GDP self-impact as well.

Table 1: Geography components participation on county distance-weighed GDP

	Average Participation				**% Growth rates**		
Component	**1930**	**1955**	**1980**	**2010**	**30-50**	**50-80**	**80-2010**
Foreign	80.66	50.80	73.59	87.58	33.44	89.98	82.27
Domestic	19.34	49.20	26.41	12.42	83.12	73.18	55.42
State	*14.59*	*34.36*	*19.32*	*9.12*	*82.21*	*74.18*	*55.33*
Neighbors	*3.11*	*11.64*	*5.41*	*2.49*	*87.14*	*69.99*	*55.56*
Local	*1.65*	*3.20*	*1.68*	*0.81*	*78.39*	*72.37*	*55.9*
TOTAL	100	100	100	100	57.89	85.52	78.89

Source: Own calculations from US Census Bureau, Internal Revenue Service for Domestic data and WTO (2005) and Maddison (2010) for international data. More information in Appendix.

One of the main issues raised in Table 1 is that the indicator has been increasing through the whole period, and most of the effect comes from the Foreign component, revealing that the commercial power of the United States has only changed recently favoring other regions. Meanwhile, county self-potential represents a small share of the domestic component.

This basic database is combined with an additional collection of data on Universities and Educational Institutions in the United States, coming from The Institutional Data Archive on American Higher Education (Brint et al. [2003]). This source contains academic data on 384 four-year colleges and universities in the United States based on stratified random sampling to over-sample elite institutions. The sample includes all highly-selective colleges and research universities in the United States, as well as other selective colleges and research universities, masters-granting comprehensive universities and non-selective baccalaureate-granting institutions. The IDA sample does not include business colleges, art schools or any other specialized institution, neither profit-institutions nor two-year program colleges. This release incorporates longitudinal and cross-sectional information on institutions, university systems, programs, academic departments, earned degrees and institutional academic rankings over time. The main variables of interest within this sample are the location of Educational Institutions, the year of establishment of these institutions and how research-intensive their activity is. This information has also been merged with the number of academic establishments provided by the County Business Patterns to gain proper visibility on the counties with no academic institutional presence.

The random relation between academic institutions is classified according to the Carnegie Classification, the leading framework for describing institutional diversity in US higher education for the past decades, (McCormick [2006]). This framework has been widely used in the study of higher education, as a way to represent and control for institutional differences and to ensure adequate representation of sampled institutions, students, or faculty. The classification has changed over time; however, the main groups are easily traceable over the period. To simplify, this database contains a numeric indicator variable from one to four, depending on the average category each institution has been over the base year in which the Classification was released. The basic categories used in our sample according to research activity includes:

- Doctorate-granting Universities: When the institution awarded at least 20 research doctoral degrees during the update year (excluding doctoral degrees that qualify professional practices, such as MD, PharmD, DPT, etc.).

- Master's Colleges and Universities: includes institutions that awarded 50 or more master's degrees and fewer than 20 doctoral degrees during the update year.

- Baccalaureate Colleges: when baccalaureate degrees represent at least 10 per cent of all undergraduate degrees and fewer than 50 master's degrees or 20 doctoral degrees were awarded during the update year.
- Associate's Colleges: when all degrees are at the associate's level or bachelor degrees account for less than 10 per cent of all undergraduate degrees.

The dataset also takes into account several control variables such as the year the county was part of the United States, whether it was one of the thirteen original colonies or the date of its official establishment as part of a State.

3.1 Empirical strategy

This paper aims to demonstrate that although income differentials are persistent over time, a human capital shock in a region will increase growth in the long run, not only in that region, but also in neighboring regions. The paper approaches this analysis by using a differences in differences regression analysis of the shock using a random sample of counties with academic institutions and a random control group of counties with no academic presence at all. These regressions compare the evolution of the distance-weighed GDP impact and address the significance of the knowledge shock in the treated group using the usual baseline specification:

$$\begin{aligned} l(GEOimpact) = \beta_0 + \beta_1(Treatment_{i,t} + \beta_2(AfterTreatment_{i,t})+ \\ +\beta_3(Treatment_{i,t} * AfterTreatment_{i,t}) + \beta_4(Controls) + e_{i,t} \end{aligned} \tag{2}$$

Where the different geographical elements are regressed on the variable *Treatment*$_{i,t}$ (which equals one if the county is in the Treated Group and zero otherwise), *After treatment*$_{i,t}$, which equals one in the period after the new university was established and the dummy interaction of both variables (*Treatment*$_{i,t}$***After treatment*$_{i,t}$), which equals one only when the county is within the Treated Group and after the treatment has taken place. Additionally, the regression controls for the different benchmark years at which the shocks take place. Using this approach, the main interest is the significance of the coefficient of the interaction term, beta$_3$, that assesses the difference of the market potential gap between the treated group and the control group after the shock has happened.

4 Results

At first glance, a quick examination of the observations with and without academic institutions samples leads to preliminary optimism regarding the potential findings from this analysis. The sample contains the 300 observations which host an academic institution in 2010 provided by the IDA database and an additional 300 random observations of counties with no academic establishment obtained from the County Business Patterns report performed by the US Census Bureau for each benchmark. As Table 2 reports, the long-term evolution of urbanization reveals that the treated group has evolved from the rural economy much faster than the control group although they started-off at very similar levels in the first year. Consistent with the consensus found by urban economists, in this sample the effects of a college education shock are positively correlated with an acceleration of city growth as proposed by Glaeser et al. [1995].

Table 2: Rural and urban counties in the sample

	Treated Group		**Control Group**	
	Rural	Urban	Rural	Urban
1930	297	3	299	1
1950	175	125	265	35
1980	121	179	235	65
2010	15	285	154	146

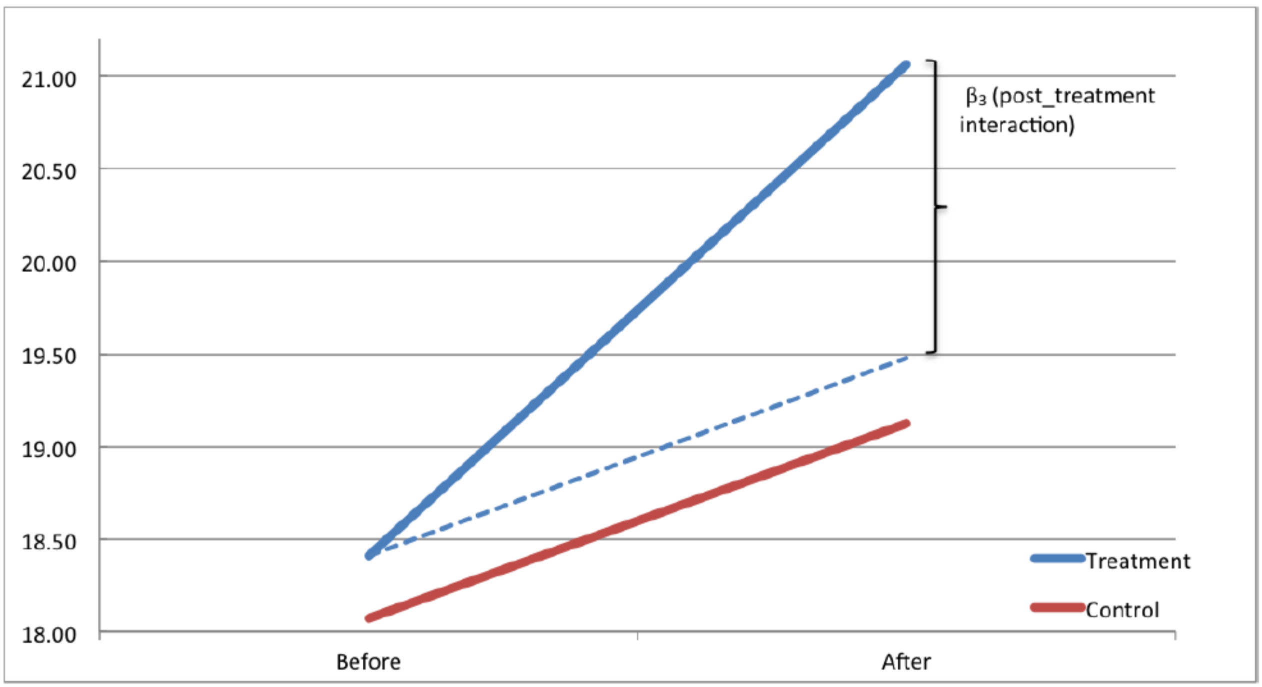
Follows the classification of Rural County provided by the US Census on each benchmark year.

The initial experiment consists of finding a causal effect between the establishment of new academic institutions and this growth differential by comparing the situation of the treated group and the control group before and after the shock. Table 4 shows the pooled OLS regressions of the geographical components of the distance-weighed sum of GDPs on the dummies related to the knowledge shock created by the new universities controlling for the year of the change. The resulting coefficients show that the shock has a positive effect on the self-potential of counties and its neighboring counties within the state at 1 per cent significance, but there is no evidence that the effect expands further. Notice the high R-squares of the regressions of the Local Impact and Neighboring County Impact in comparison with the rest of the components; other unobserved variables affect GDP of States and Foreign competitors.

The significance of the shock on the local element is not unexpected, as it is consistent with the central premises of education economics. Many have argued for the simultaneous causality of this result though, and have even posed education as the 'weak link' of the growth literature: higher local GDP (the central element of the indicator of the Local element of the distance-weighed GDP) fosters a higher provision of services and vice-versa.[3] Proving the causality behind such a relationship is beyond the scope of this investigation that relies on the findings of scholars like Goldin and Katz [2009],Moretti [2002], and Rauch [1993].

The most interesting result comes from the positive and significant coefficient of the interaction term in the Neighboring Counties regression. The neighboring counties GDP component might be correlated with the local GDP component, however it is bound to be affected by many other independent variables. Thus, controlling for yearly fixed effects and regional effects adds reliability to the results of the neighbors impact regression. The effect of the shock reports a 7 per cent increase of the impact on neighboring counties in the Treated Group than in the Control Group at one per cent significance, as illustrated in Figure 1.

Figure 1: Effect of the new university on neighboring counties with Random Control Group



Source: from own calculations.

These results yield evidence of the significance of the treatment on nearby areas. However, one might question these results by addressing the self-selection of counties to the location of new universities even by abstracting from the obviously endogenous character

[3] Krueger and Lindahl [2001]; Mankiw [1997].

of self-potential. It is possible that new universities might be allocated in areas where high growth is already expected. The decision to establish a new academic institution might not even be taken locally but at a more global scale as in the case of the Morrill Land-Grant Acts (1862), where federal incentives motivated the creation of state universities by state governments. State governments could have selected these allocations in areas with positive growth prospects.

This intuition leads to the hypothesis that the Treated Group counties might be special. In order to control for this potential self-selection, an alternative synthetic random control group that shows similar initial characteristics to the ones in the treated group could be used to perform the same analysis, replicating the technique developed by Abadie and Gardeazabal [2003]. This way, the difference-in-difference coefficients will only show the variance related to the treatment, eliminating the bias from the intrinsic heterogeneity of the samples.

4.1 Synthetic control method

Finding a control group that is synthetically equivalent to the Treatment Group implies restricting the sample to comply with several conditions that are met by the treatment group at the beginning of the period. In this sense, the Synthetic Control Group must start-off at a level similar to the treated group. Indeed, from Table 3 it is visible that the Treatment Group is a special sample of counties and rather different to the the Random Control Group used in the previous section. An examination of the initial levels of urbanization was not a sufficient way to analyze the impact of new academic institutions.

The Treated Group does not only differ substantially from the average level of primary sector employment of the population, but is almost half of that of the Random Control Group. The shares devoted to other sectors consequently also differ, thus showing that our counties in both samples evidence very different productive structures and pools of labor. Additionally, the pattern of growth seems to have diverged in both samples, where population from the Control Group seems to be stagnant between 1930 and 1980 while the counties in the Treated Group have increased their population by almost 50 per cent in the same period. It seems, thus, that the previous analysis considers two rather contrasting samples of counties, where clearly most of the components of the weighed GDP sum amount to different levels. Consequently, the alternative random control group has been forced to meet certain criteria, such as a similar productive structure or a closer local component sum of distance-weighed Total GDP, finding an alternative sample of 300 counties with no academic presence with a similar productive structure and distance-weighed Total GDP.

In line with the previous case, note that the initial level of the Synthetic Random Group starts-off at slightly higher levels of urbanization than the treated group with an even smaller primary sector participation share, but the rate of urbanization is quite slower than that of the Treated Group with slower population growth from 1930 to 1980. There is still hope to find optimistic results regarding the shock.

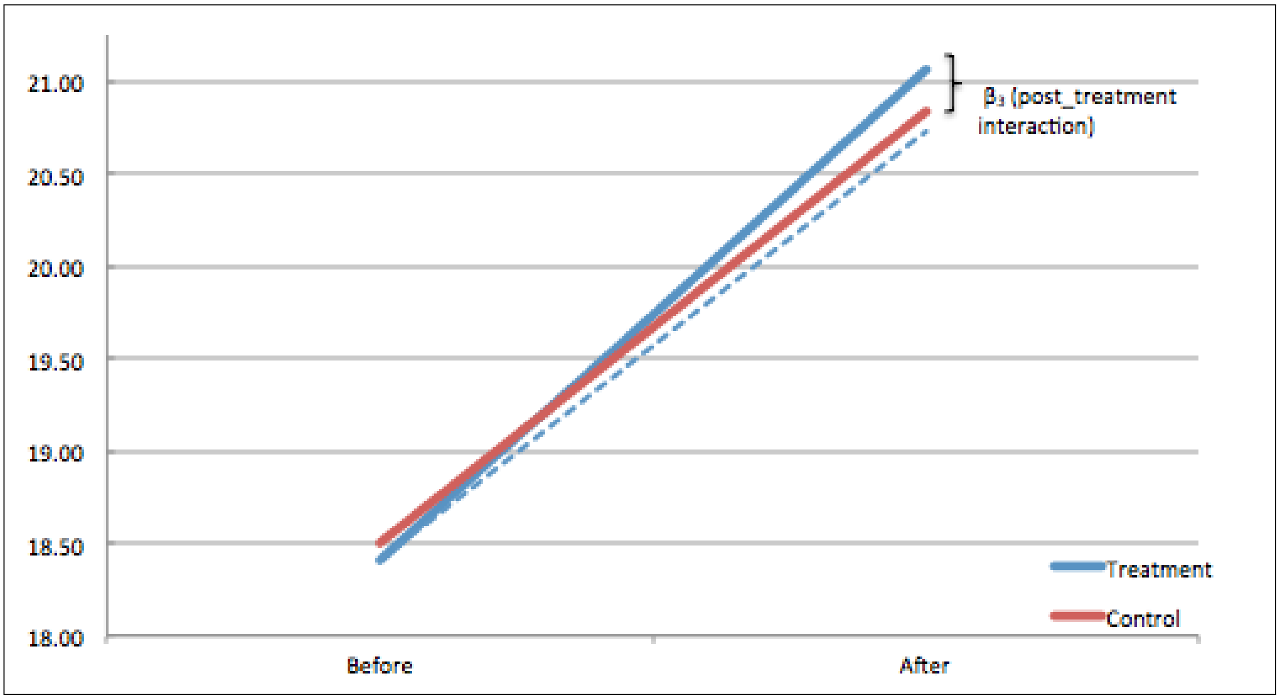
Table 3: Descriptive data from samples and population

Group	All counties	Treatment	Random	Synthetic
Establishments		=1	=0	=0
	N=3128	N=300	N=300	N=300
Average Indicators in 1930	T=4	T=4	T=4	T=4
l(Distance-weighed Total GDP)	18.284	19.154	17.942	19.395
l(Local component)	14.347	14.865	14.237	14.903
l(Neighbor Cty. component)	16.073	16.544	15.955	16.691
l(States component)	15.911	15.708	15.949	16.236
l(Foreign component)	16.277	17.709	15.601	17.707
Employment Share %				
Primary Sector	0.487	0.289	0.578	0.279
Secondary Sector	0.132	0.256	0.068	0.259
Tertiary Sector	0.380	0.456	0.354	0.461
Population growth (1930-1980)	0.113	0.450	-0.051	0.382
Rural counties in sample				
1930	3117 (0.99)	297 (0.99)	299 (0.99)	300 (1.00)
1950	2844 (0.91)	175 (0.58)	265 (0.88)	279 (0.93)
1980	2462 (0.79)	121 (0.40)	235 (0.78)	234 (0.78)
2010	1325 (0.43)	15 (0.05)	154 (0.51)	67 (0.22)

Source: own calculations.

The analysis of the treatment with the Synthetic Control Group provides a better view of the treatment with unbiased estimators of the interaction group. This time, regression results show coherent signs and sizes and are significant in all the geographical components of the impact. According to Table 5, the significance and size of the local impact has decreased by a third, but the impact of the shock is now relevant in terms of counties, states and foreign counties. In other words, after a new university is established in Fresno, all the counties in California will experience an upswing in their GDP (*ceteris paribus*) significant at 1 per cent. Moreover, this shock will have a positive statistical impact on the rest of the counties within the country, although the effect will be smaller than for the neighbors within the same state, as opposed to the previous experiment with the Random Control Group. Additionally, the shock seems to make the domestic economy more competitive, creating a negative and quite sizeable Foreign GDP impact.

Figure 2: Effect of the new university on neighboring counties with Synthetic Control Group



Source: from own calculations.

The knowledge shock, thus, seems quite important to future development both in terms of regional and global inequality; however, one could question whether quality could also be an issue. To control for quality, the same analysis has been repeated including the interaction of a variable that accounts for the average Carnegie classification rate of each institution to show whether more research prone institutions have a higher impact than associate college institutions. Results show no significant evidence on the difference associated to the level of research-intensity; in other words, there is no evidence of a higher effect of doctorate-granting academic institutions over Baccalaureate colleges; instead, any academic institution that creates a human capital shock will have a significant impact.

To improve long-term visibility, the same regression can be adapted to include several time periods to account for the date of the shock, where the independent variable is the final-benchmark year distance-weighed GDP impact. Table 6 shows the differences in differences regression of the long term impact of each shock. This table provides a much more detailed view of the shock that confirms that the local impact expands to nearby regions, diluting its effect with distance, while also having a significant effect in the international arena.

Table 4: Before & after effect of new university against Random Control Group

Impact	l(Total)		l(Local)		l(Neighbors)		l(States)		l(Foreign)	
Treatment group	1.619	***	0.311	**	0.336		0.003		3.277	***
	(0.344)		*(0.161)*		*(0.229)*		*(0.603)*		*(0.435)*	
After Treatment	-1.267	***	-1.837	***	-1.773	***	-1.739	***	-0.203	
	(0.294)		*(0.094)*		*(0.126)*		*(0.072)*		*(0.171)*	
PostTreatInteraction	0.230		0.318	***	0.226	**	0.108		-0.739	***
	(0.265)		*(0.132)*		*(0.114)*		*(0.069)*		*(0.218)*	
N	1384		1380		1372		1384		1384	
r-squared	0.396		0.490		0.629		0.304		0.408	

Coefficients from the OLS difference-in-differences regression of the log of the geographical components of distance-weighed sum of GDPs to the shock of new universities. Robust standard errors clustered by state are reported under the coefficients. *** indicates statistical significance at 1%, ** at 5% and * at 1%. *Source: own calculations.*

Table 5: Before & after effect of new university against Synthetic Control Group

Impact	l(Total)		l(Local)		l(Neighbors)		l(States)		l(Foreign)	
Treatment group	0.609	***	-0.100		-0.094		0.087		1.492	***
	(0.286)		*(0.196)*		*(0.178)*		*(0.494)*		*(0.483)*	
After Treatment	0.126		-1.532	***	-1.425	***	-1.313	***	1.145	***
	(0.219)		*(0.087)*		*(0.102)*		*(0.110)*		*(0.334)*	
PostTreatInt	-0.854	***	0.276	**	0.317	***	0.142	***	-1.756	***
	(0.186)		*(0.106)*		*(0.084)*		*(0.052)*		*(0.353)*	
N	1388		1388		1388		1388		1388	
R-squared	0.445		0.441		0.587		0,275		0,315	

Coefficients from the OLS difference in difference regression of the log of the geographical components of distance-weighed sum of GDPs to the shock of new universities. Robust standard errors clustered by state are reported under the coefficients. *** indicates statistical significance at 1%, ** at 5% and * at 1%. *Source: own calculations.*

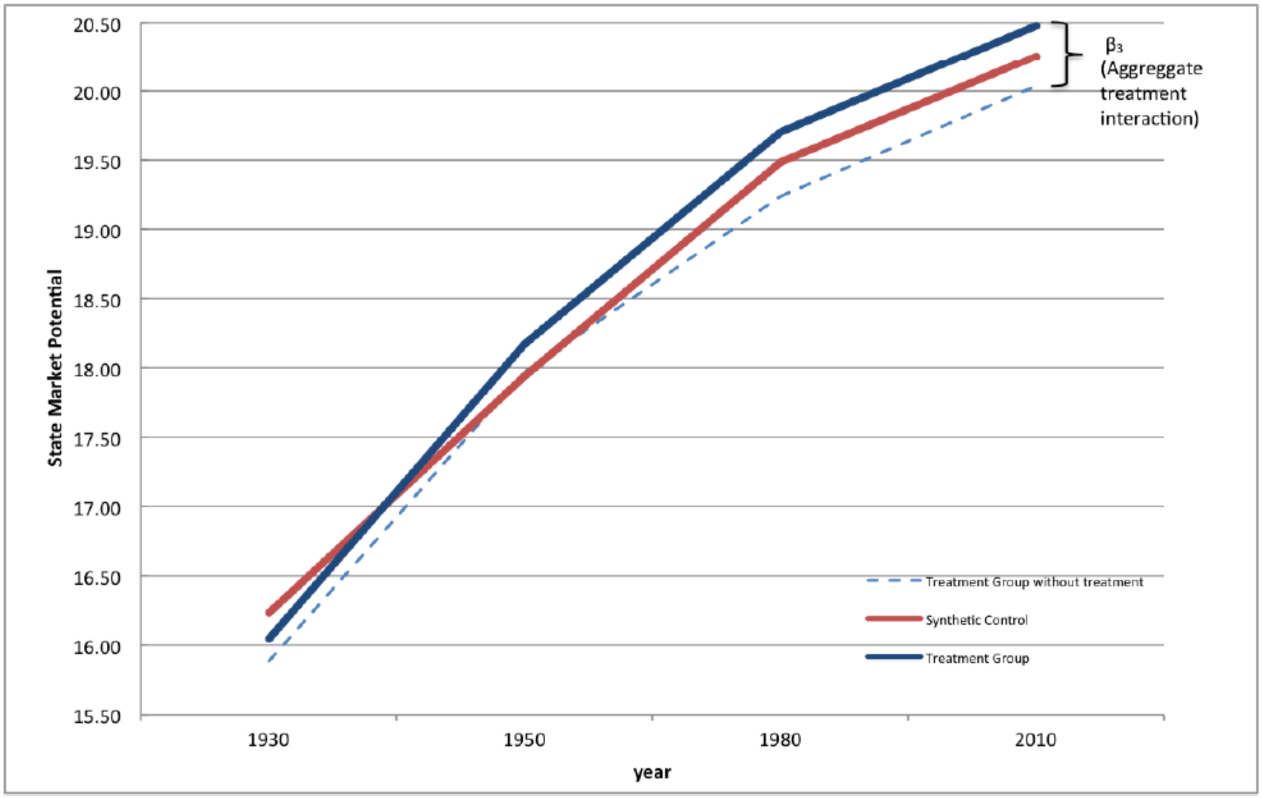
The effect of time seems to be similar: on average, the closer in time the shock is, the higher the effect becomes, although there seem to be discrepancies in the long run impacts of the different geographical layers: the domestic impact seems to last longer than the foreign impact according to the significance of the interaction treatment variables.

This new experiment shows that, for counties with similar levels of urbanization, a knowledge shock implies a regional acceleration of GDP growth that expands to nearby regions, creating a 'shock-wave' effect that also impacts faraway counties. Although the Synthetic Control Group seems to start slightly above the Treated control group in terms of distance-weighed Total GPD, the treatment has led the Treated Group to surpass the Synthetic Control Group by a lower, but still significant, impact as shown in Figure 2. Overall, the shock affects the relative position of the whole country in international perspective. This impact is independent of the type of institution that creates the knowledge shock. However, this before-after treatment analysis does not say much about the long-term effects of the shock and whether the change is persistent or temporary.

A possible interpretation of the time coefficients can come from the statistical significance of single time variables. It seems that time affects significantly the evolution of regional GDP weighed by economic distance. Firstly, transport costs decrease over time (see the Appendix for more details), and so does economic distance leading to an overall increase of the distance-weighed sum of GDPs. Second, path-dependency is rather crucial when determining income; initial GDP is a rather important factor of future GDP. Consequently, GDP ten years ago also matters, but to a lesser degree. According to Table 6, the impact of the shock has shorter term consequences for the local economy than for neighboring counties or states. If the treatment took place in the 1940s and showed its consequences in the 1950s, it is natural that the effect of the treatment is already taken into account in the time variable rather than the interaction variable. However, the interaction has longer term consequences for neighboring areas than for the local impact. This result may seem unreasonable, however I propose an interpretation related to increasing returns to scale. The local shock of a new university institution might be locally absorbed almost immediately, or in a few years, however, spillover effects may take some time to reach neighboring counties, and even more if they reach further. In practical terms, new researchers rapidly arrived to the Idaho Falls city as soon as the job positions started, but the expansion of the city took more time.

Similarly, the long-term interaction of the Foreign impact component might require some explanation. In this case, it seems to show a similar behavior to the Local impact, evidencing a single short term impact. Notice that the Foreign impact is assembled by the GDP of countries weighed by the economic distance (bilateral transport costs from each county to each county). Each country's GDP is affected by many more variables that are independent of the events in the USA, including the response to the opening of new academic institutions, which might have an analogue negative effect in US counties.

Figure 3: Long term effect of the new university on neighboring counties with Synthetic Control Group



Source: from own calculations.

Table 6: Effect of new university against Synthetic Control Group in the long term

Impact	**l(Total)**		**l(Local)**		**l(Neighbors)**		**l(States)**		**l(Foreign)**	
Treatment group	0.461	***	-0.198		-0.312		-0.187		1.348	***
	(0.317)		*(0.226)*		*(0.218)*		*(0.500)*		*(0.514)*	
Before 2010	4.502	***	3.644	***	4.150	***	4.015	***	5.206	***
	(0.216)		*(0.141)*		*(0.186)*		*(0.178)*		*(0.340)*	
Before 1980	2.718	***	3.158	***	3.354	***	3.247	***	2.642	***
	(0.166)		*(0.123)*		*(0.165)*		*(0.180)*		*(0.206)*	
Before 1950	1.030	***	1.758	***	2.085	***	1.706	***	0.086	***
	(0.171)		*(0.113)*		*(0.109)*		*(0.172)*		*(0.208)*	
PostTreat2010	-0.707		0.373	**	0.535	***	0.416	***	-1.612	***
	(0.243)		*(0.174)*		*(0.139)*		*(0.150)*		*(0.390)*	
PostTreat1980	0.253		0.196		0.456	***	0.404	***	0.125	
	(0.163)		*(0.155)*		*(0.135)*		*(0.147)*		*(0.217)*	
PostTreat1950	0.190	***	0.095		0.198	**	0.417	***	0.310	
	(0.185)		*(0.129)*		*(0.098)*		*(0.147)*		*(0.229)*	
Constant	19.395	***	14.890	***	16.199	***	1.236	***	17.687	***
	(0.289)		*(0.234)*		*(0.533)*		*(0.535)*		*(0.449)*	
N	1388		1388		1388		1388		1388	
R-squared	0.445		0.450		0.609		0.279		0.325	

Coefficient from the difference in difference regression of the log of the geographical components of distance-weighed Total GDP sum to the shock of new universities. Robust standard errors clustered by state are reported under the coefficients. *** indicates statistical significance at 1%, ** at 5% and * at 1%. *Source: own calculations.*

4.2 Local endogeneity test

Previous sections acknowledged the potential endogeneity of the local element, exposing the simultaneous causation between higher local GDP and higher provisions of services like university schooling posited by Mankiw [1997] and Krueger and Lindahl [2001]. Although proving the causality behind such a relationship is beyond the scope of this investigation, the variables used in this analysis allow for the performance of a simple extension that clarifies the link between local productivity and these knowledge shocks.

A slight transformation of the distance-weighed Total GDP sum to per capita terms has two consequences: first, it allows for the removal of the effect of the size of each of the participants in the sum of GDPs, while making distance much more important. Secondly, it allows one to understand the impact of a new knowledge institution in terms of productivity (as per capita GDP is a proxy for wages).

As a result, the new independent variable is the sum of the distance-weighed sum of per capita GDPs, where the foreign impact becomes relatively smaller than the domestic components. This is because the bigger size of countries is controlled for while the effect of its further distance becomes much more important. Likewise, the impact of close neighboring counties increases over distant counties, states and countries. Because the local component is computed as each county's GDP per capita over the great circle distance equivalent to its area, high wages and small areas lead to bigger local effects in the sum of the distance-weighed sum of per capita GDPs. In other words, big city counties with high productivities have relatively higher local components than large rural counties; furthermore, the total sum of distance-weighed per capita GDPs is driven by the local component as can be seen in Figure 5 in the Appendix.

Table 7 shows the result of performing the parallel regression between the Treatment group and the Synthetic Control Group in per capita terms. This time the results are quite distinct, leading to subtle differences between the growth of the economy and its development through improvements in labor quality. First, the effect of the shock on the State and Foreign components is no longer significant: the local shock does not affect distant economies. However, it seems that the local effect of the shock has now increased its size and significance and also affects nearby regions.

Table 7: Effect of new university against Synthetic Control Group per capita in the long term

Per capital impact	**l(Total)**	**l(Local)**	**l(Neighbors)**	**l(States)**	**l(Foreign)**
Treatment group	0.245	-0.278	0.076	0.255	0.730 *
	(0.177)	*(0.222)*	*(0.561)*	*(0.169)*	*(0.447)*
Before_2010	-0.088	3.644 ***	-0.860	0.032	0.078
	(0.124)	*(0.141)*	*(0.306)*	*(0.113)*	*(0.288)*
Before_1980	0.007	2.934 ***	-0.539	0.098	0.029
	(0.177)	*(0.192)*	*(0.352)*	*(0.163)*	*(0.354)*
Before_1950	-0.109	1.600 ***	-0.368	-0.113	-0.360
	(0.206)	*(0.162)*	*(0.361)*	*(0.203)*	*(0.337)*
Post_treatment2010	0.088	0.448 ***	0.860 ***	-0.032	-0.078
	(0.124)	*(0.190)*	*(0.306)*	*(0.113)*	*(0.288)*
Post_treatment1980	-0.069	0.421 **	0.539	-0.098	-0.029
	(0.177)	*(0.206)*	*(0.352)*	*(0.163)*	*(0.354)*
Post_treatment1950	0.109	0.253	0.368	0.113	0.360
	(0.206)	*(0.156)*	*(0.361)*	*(0.203)*	*(0.337)*
Constant	20.448	14.890 ***	17.536 ***	20.099 ***	6.340 ***
	(0.212)	*(0.234)*	*(0.581)*	*(0.180)*	*(0.269)*
N	1388	1388	1388	1388	1388
R-squared	0.011	0.510	0.014	0.009	0.032

Coefficient from the difference in difference regression of per capita geographical components of distance-weighed Total GDP sum to the shock of new universities. Robust standard errors clustered by state are reported under the coefficients. *** indicates statistical significance at 1%, ** at 5% and * at 1%. *Source: own calculations.*

Additionally, the effect of a local shock does not only increase the local GDP per capita in the first period after the shock, but persists after one period on a lower scale, whereas the effect on the productivity of nearby regions is constrained to the first period of the shock.

These results show that knowledge shocks do affect the local economy as well as nearby regions. By controlling the potential causality between increasing counties and increasing service provision, this experiment has shown that the effect of knowledge shocks go beyond the size of the economy and affect the local productivity of regions and their neighbors, leading to multiplier effects that explain the results in Table 6.

5 Conclusions

This paper has offered an alternative methodology to explore the regional impact of local shocks from different geographical perspectives. Traditionally, academics have used spatial econometrics frameworks that require the application of tools like neighboring matrices that limit the study to small or aggregated areas. Using this methodology allows a more detailed analysis that ensures unbiased results by including the whole population. Using a traditional differences in differences regression analysis on a decomposable sum of distance-weighed GDP by county, this study has shown that the impact of a new university affects not only the local economy, but also other counties within the state. This was already proven by Anselin et al. [1997] in the context of manufacturing industries. The analysis of the distance-weighted Total GDP reveals that the impact also affects counties in other states and improves the relative international competitiveness of the country. Using a Synthetic Control Group that replicates the initial conditions of the Treated Group and comparing their evolution after the treatment, I show that the effect of the shock is not spurious and, although smaller, it is still significant.

The effect of the shock dilutes with distance and time but remains significant although the time impact of the different geographical components varies, being reduced for the foreign element that obviously depends on many other factors, like its own domestic policy. The regional effect of the shock seems to take some time but is persistent in the long run, *ceteris paribus*.

While land abundance was originally posed as a curse for Idaho Falls, the establishment of a research center transformed the production frontier of this desert area to a more valuable bundle of products that included knowledge intensive services. The previous results and this anecdotal evidence provide a lesson for both local and national authorities: economic fundamentals are by no means a restriction to the production possibility frontier; any economic configuration can provide increasing returns to scale. Perhaps, further research could help develop a model to find the factors that define the optimum industry for maximizing both the local impact and spillovers. This might require some effort from academics and policy-makers, but this strategy could both reduce inequality and perpetuate the USA as the human capital paradigm.

References

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, pages 113–132.

Anselin, L. (2000). Geographical spillovers and university research: A spatial econometric perspective. *Growth and Change*, 31(4):501–515.

Anselin, L., Varga, A., and Acs, Z. (1997). Local geographic spillovers between university research and high technology innovations. *Journal of Urban Economics*, 42(3):422–448.

Atack, J., Bateman, F., Haines, M., and Margo, R. A. (2010). Did railroads induce or follow economic growth? *Social Science History*, 34(2):171–97.

Audretsch, D. B. and Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review*, pages 630–640.

Audretsch, D. B., Hülsbeck, M., and Lehmann, E. E. (2012). Regional competitiveness, university spillovers, and entrepreneurial activity. *Small Business Economics*, 39(3):587–601.

Autor, D., Katz, L., and Kearney, M. (2008). Trends in US wage inequality: Revising the revisionists. *The Review of Economics and Statistics*, 90(2):300–323.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, pages 775–805.

Brakman, S., Garretsen, H., and Schramm, M. (2004). The strategic bombing of German cities during World War ii and its impact on city growth. *Journal of Economic Geography*, 4(2):201–218.

Brint, S. L., Riddle, M., and Levy, C. S. (2003). The Institutional Data Archive on American Higher Education.

Caselli, F. and Coleman, W. J. (2002). The US technology frontier. *American Economic Review*, pages 148–152.

Ciarli, T., Meliciani, V., and Savona, M. (2012). Knowledge dynamics, structural change and the geography of Business Services. *Journal of Economic Surveys*, 26(3):445–467.

Crafts, N. and Mulatu, A. (2005). What explains the location of industry in Britain, 1871–1931? *Journal of Economic Geography*, 5(4):499–518.

Davis, D. R. and Weinstein, D. E. (2002). Bones, bombs, and break points: The geography of economic activity. *The American Economic Review*, 92(5):1269–1289.

Drucker, J. and Goldstein, H. (2007). Assessing the regional economic development impacts of universities: a review of current approaches. *International Regional Science Review*, 30(1):20–46.

Edwards, A. M. (1943). Comparative occupation statistics for the United States, 1870 to 1940.

Enflo, K. and Rosés, J. R. (2015). Coping with regional inequality in Sweden: structural change, migrations, and policy, 1860–2000. *The Economic History Review*, 68(1):191–217.

Fujita, M. and Krugman, P. (2003). The new economic geography: Past, present and the future. *Papers in Regional Science*, 83(1):139–164.

Geiger, R. L. (2014). *The History of American Higher Education: Learning and Culture from the Founding to World War II*. Princeton University Press.

Glaeser, E. L., Ponzetto, G. A., and Tobio, K. (2014). Cities, skills and regional change. *Regional Studies*, 48(1):7–43.

Glaeser, E. L., Scheinkman, J., and Shleifer, A. (1995). Economic growth in a cross-section of cities. *Journal of Monetary Economics*, 36(1):117–143.

Goldin, C. D. and Katz, L. F. (2009). *The Race Between Education and Technology*. Harvard University Press.

Golub, S. S. and Tomasik, B. (2008). Measures of international transport cost for OECD countries.

Grilliches, Z. (1979). Issues in assessing the contribution of R+D to productivity growth. *Bell Journal of Economics*, 100.

Hanlon, W. W. (2014). Temporary shocks and persistent effects in the urban system: Evidence from British cities after the US Civil war. *NBER Working Paper*, (w20471).

Harris, C. D. (1954). The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers*, 44(4):315–348.

Hornbeck, R. and Keniston, D. (2014). Creative destruction: Barriers to urban growth and the Great Boston fire of 1872. *NBER Working Paper*, (w20467).

Jacks, D. S., Meissner, C. M., and Novy, D. (2008). Trade costs, 1870-2000. *The American Economic Review*, pages 529–534.

Jaffe, A. B. (1989). Real effects of academic research. *The American Economic Review*, pages 957–970.

Krueger, A. B. and Lindahl, M. (2001). Education for growth: Why and for whom? *Journal of Economic Literature*, 39:1101–1136.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42.

Mankiw, N. G. (1997). The neoclassical revival in growth economics: Has it gone too far?: Comment. *NBER Macroeconomics Annual*, pages 103–107.

Martínez-Galarraga, J. (2012). The determinants of industrial location in Spain, 1856–1929. *Explorations in Economic History*, 49(2):255–275.

McCormick, A. C. (2006). The Carnegie Classification of Institutions of Higher Education.

Missiaia, A. (2014). *Industrial location, market access and economic development: regional patterns in post-unification Italy*. PhD thesis, London School of Economics and Political Science (LSE).

Mohammed, S. I. S. and Williamson, J. G. (2004). Freight rates and productivity gains in British tramp shipping 1869–1950. *Explorations in Economic History*, 41(2):172–203.

Moretti, E. (2002). Human capital spillovers in manufacturing: Evidence from plant-level production functions. Technical report, US Census Bureau, Centre for Economic Studies.

Moretti, E. (2004). Human Capital Externalities in Cities. *Handbook of Regional and Urban Economics*, 4.

Moretti, E. (2010). Local multipliers. *The American Economic Review*, 100(2):373–377.

Moretti, E. (2012). *The New Geography of Jobs*. Houghton Mifflin Hartcourt Publishing Company, 2013 edition.

Rauch, J. E. (1993). Productivity gains from geographic concentration of human capital: evidence from the cities. *Journal of Urban Economics*, 34(3):380–400.

Redding, S. J., Sturm, D. M., and Wolf, N. (2011). History and industry location: Evidence from German airports. *Review of Economics and Statistics*, 93(3):814–831.

Simon, C. J. and Nardinelli, C. (2002). Human capital and the rise of American cities, 1900–1990. *Regional Science and Urban Economics*, 32(1):59–96.

Primary Sources

- **Internal Revenue Service** (1930): *Statistics of Income, Table 6.*
- **Internal Revenue Service** (1950): *Statistics of Income, Table 13.*
- **Mayer, Thierry and Zignago, Soledad** (2006): *CEPII - "GeoDist"*.
- **NOAA** (2010): *US National Oceanic and Atmospheric Administration - Department of Commerce*.

- **US Census Bureau** (2010): *American Community Survey*.
- **US Census Bureau** (2010): *Housing and Population Censuses*. Decennial Records.
- **US Census Bureau** (2010): *Topologically Integrated Geographic Encoding and Reference System*.
- **US National Oceanic and Atmospheric Administration** (2010): *Department of Commerce*.
- **World Trade Organisation** (2005): *International Trade Statistics*.

Appendix

Sources and Methods for distance-weighed sum of GDP by county calculation

This calculation is based on the original definition of Market Potential that was proposed by Harris [1954] as an indicator of a location's accessibility to other markets formulated as the sum of the rest of the regions size (measured by GDP) and weighed by bilateral transport costs or distances. Each county-year observation of the distance-weighed sum of GDP has been calculated following Crafts and Mulatu [2005] and Martínez-Galarraga [2012], who use the traditional definition of MP as the sum of the size of potential markets measured by GDP weighed by the bilateral economic distances, as follows:

$$MP_i = \sum_{(j-1)}^{(j-n)} \frac{M_j}{d_i j} \tag{3}$$

Economic distance is the distance between pairs weighted by transportation costs. These have been estimated based in Jacks et al. [2008], Mohammed and Williamson [2004] and Harris [1954] data and are shown in Table 8 below. Total MP is further decomposable by the domestic and foreign effect:

$$MP_i = DMP_i + FMP_i \tag{4}$$

$$MP_i = \sum_{(j-i)}^{(1,n)} \frac{M_s}{d_i, s} + \sum_{(US-i)}^{(1,n)} \frac{M_U S}{d_i, US} + SP_i + FMP_i \tag{5}$$

In this dataset, counties distance-weighed total GDP has been decomposed by its Domestic Impact and Foreign Impact, in parallel with the MP formula. More than this, the domestic element can be further disentangled as the State's impact, the within state neighboring counties impact and *i*'s own local impact. Following Crafts and Mulatu [2005] and Martínez-Galarraga [2012], local component has been found as county's size measured by GDP, divided by the radius of the circle with an equivalent area of the county to control for distance:

$$MP_i = \sum_{(j-i)}^{(1,n)} \frac{M_s}{d_i, s} + \sum_{(US-i)}^{(1,n)} \frac{M_U S}{d_i, US} + \frac{\frac{E_i}{E_s}(Ms)}{(1/3)\sqrt{\frac{area_c}{\pi}}} + \sum_{(j-i)}^{(1,n)} \frac{M_F}{d_i, F} \cdot d_c, F^{-0.8} \cdot coast_c \tag{6}$$

Economic distance-weighed GDP between each county and its international trading partners has been obtained by designating domestic nodes just as Market Potential calculations are usually done (Martínez-Galarraga [2012] and Missiaia [2014]). In this case, the procedure is closer to the one followed by Jacks et al. [2008], by choosing closest distance from each county to any of the top 100 biggest cities from 1930 to 1980 or the Standard Statistical Metropolitan Areas in 2010 as the most probable connection of trade.

The 30 countries used as trading partners of the US are based on the top importers on each benchmark years reported by the World Trade Organisation [2005] and on the historical data availability on GDP obtained from Maddison [2010].[4] GeoDist (Mayer and Zignago [2006]), provide information on the geodesic bilateral distances from counties to each of countries that are used as commercial partners of the US. The latitudes and longitudes of the centroid of each polygon on all US counties and its area are available from and US Census Bureau. A dummy variable for county coasts was built to account for international transportation costs and was obtained following the criteria of the National Oceanic and Atmosphere Administration of the USA Department of Commerce. Bilateral distances have been obtained from latitude and longitude coordinates available from TIGER (US Census Bureau).

GDP by state for the years 2010, 1980, 1950 and 1930 have been obtained from the original records of the Internal Revenue Service, that hold the original documents from Statistics of Income reports for 1930, 1950 and 1980 and from the Bureau of Economic Analysis for 2010. These original data were presented in nominal dollars and had to be transformed to real dollars using CPI deflator's calculations from Bureau of Labour Statistics.

Additionally, the transport costs structure has been derived from the information presented in several sources. International transport costs have been obtained from the long-term evolution presented in Mohammed and Williamson [2004] and Jacks et al. [2008], the average bilateral costs presented by Golub and Tomasik [2008] for bilateral trade costs between US and OECD countries. Internal transport costs have been obtained from the data presented by Harris [1954] and the Statistical Abstracts for each year's economic census. The cost structure has been calculated as an ad-valorem tariff equivalent that follows the following structure:[5]

Table 8: Trade Costs

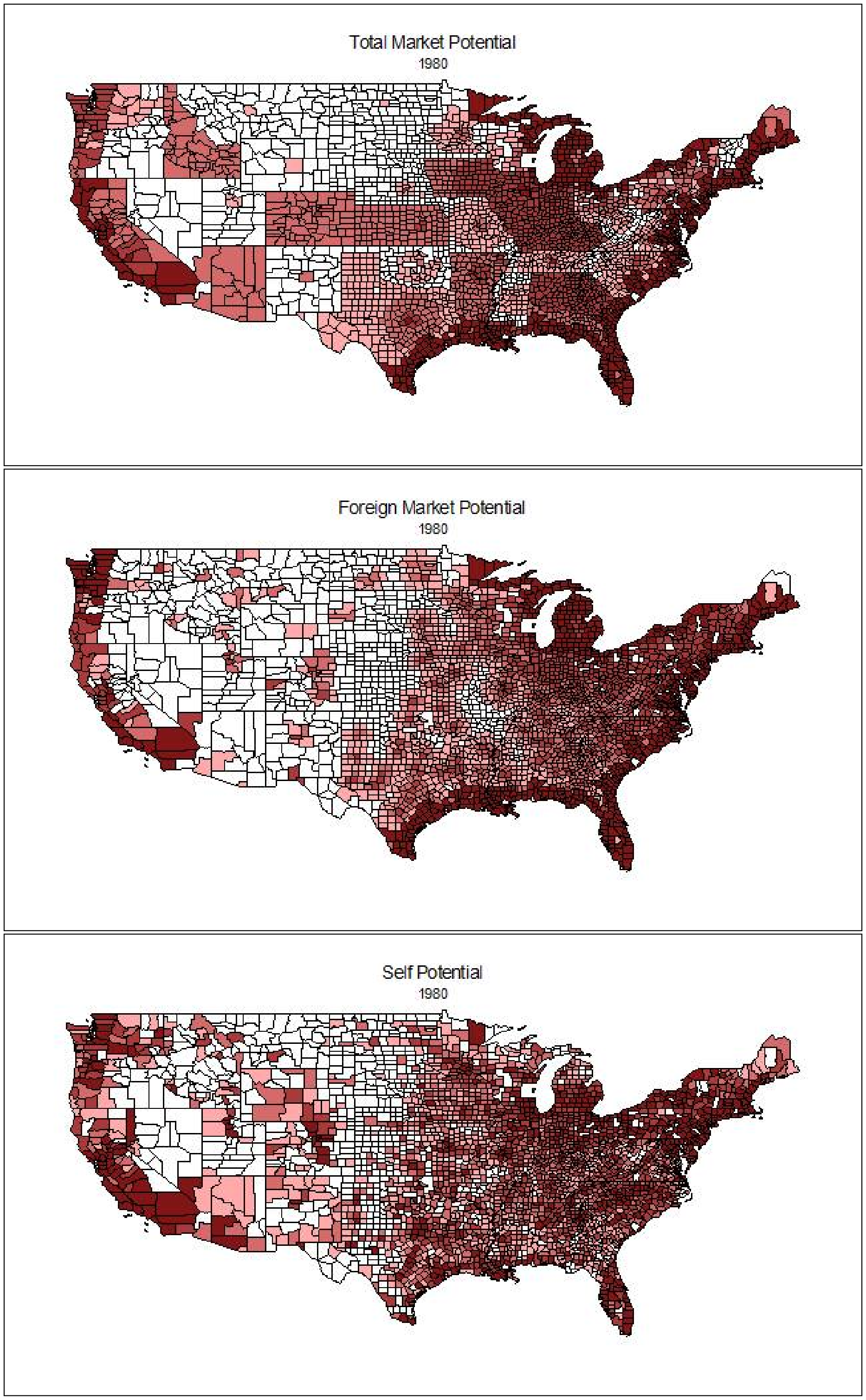
	1930	**1950**	**1980**	**2010**
INTERIOR				
Trucking				
0-80 Kms	0.037	0.029	0.017	0.004
80-160 Kms	0.049	0.039	0.023	0.006
160-480 Kms	0.061	0.049	0.029	0.007
480-708 Kms	0.109	0.087	0.052	0.013
Railroads				
708-1780 Kms	0.139	0.136	0.119	0.102
more than 1780 Kms	0.215	0.223	0.196	0.168
INTERNATIONAL	0.292	0.365	0.156	0.146

Ad-valorem tax-equivalent per potential transported dollar. *Source: own calculations.*

[4] Argentina, Australia, Belgium, Brazil, Canada, Colombia, Cuba, Denmark, Finland, France, Germany, Greece, India, Ireland, Italy, Japan, Malaysia, Mexico, Netherlands, Philippines, Poland, Portugal, Spain, Switzerland, Taiwan, United Kingdom, Venezuela and, since 1950, Israel, Singapore and South Korea.

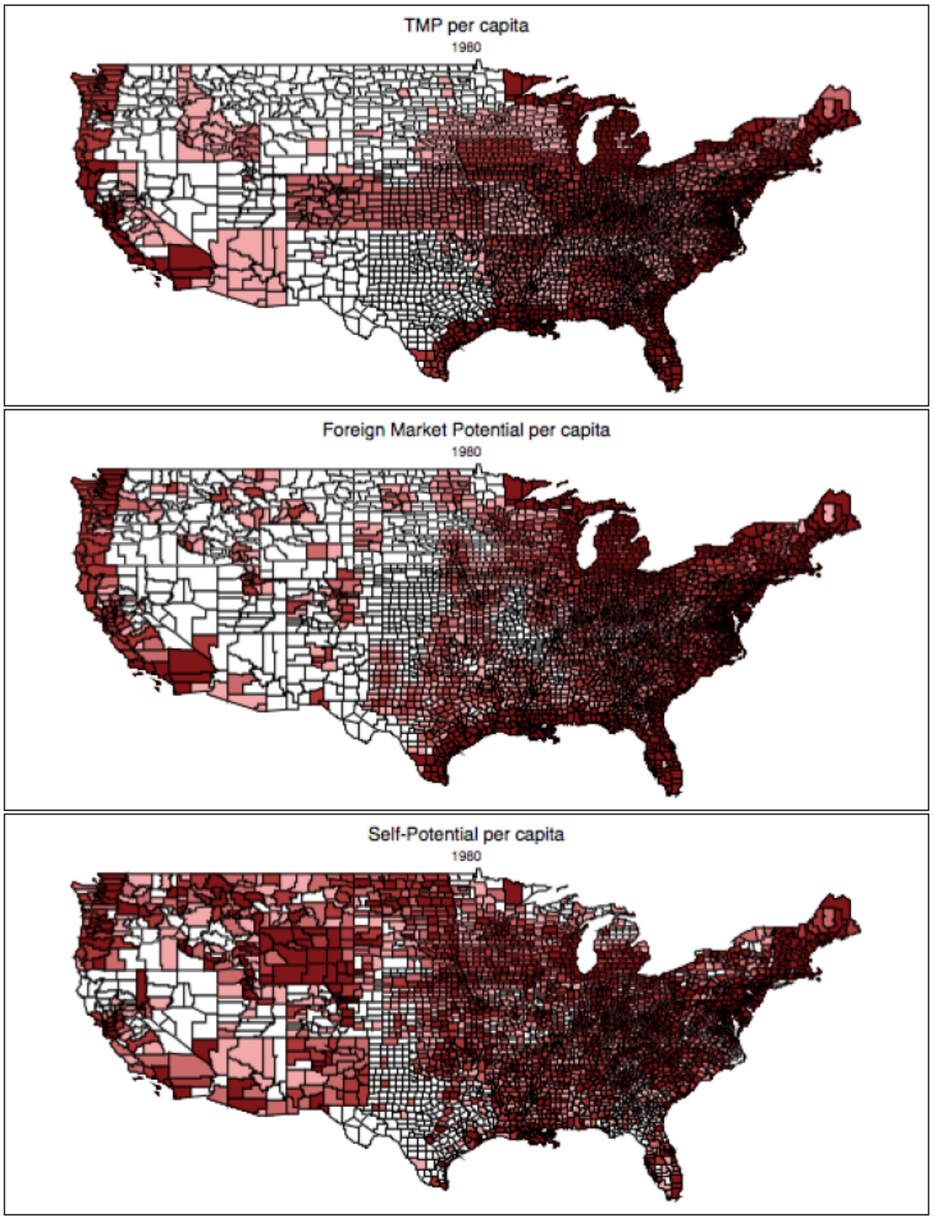
[5] Golub and Tomasik [2008] show that the cost of transportation by kilogram of goods transported is very similar to the cost of transportation by dollar.

Figure 4: Distance-weighed sum of GDP components, 1980



Source: from author's own calculations.

Figure 5: Distance-weighed sum of per capita GDP components, 1980



Source: from author's own calculations.